

CELEBAL TECHNOLOGIES PVT. LTD.

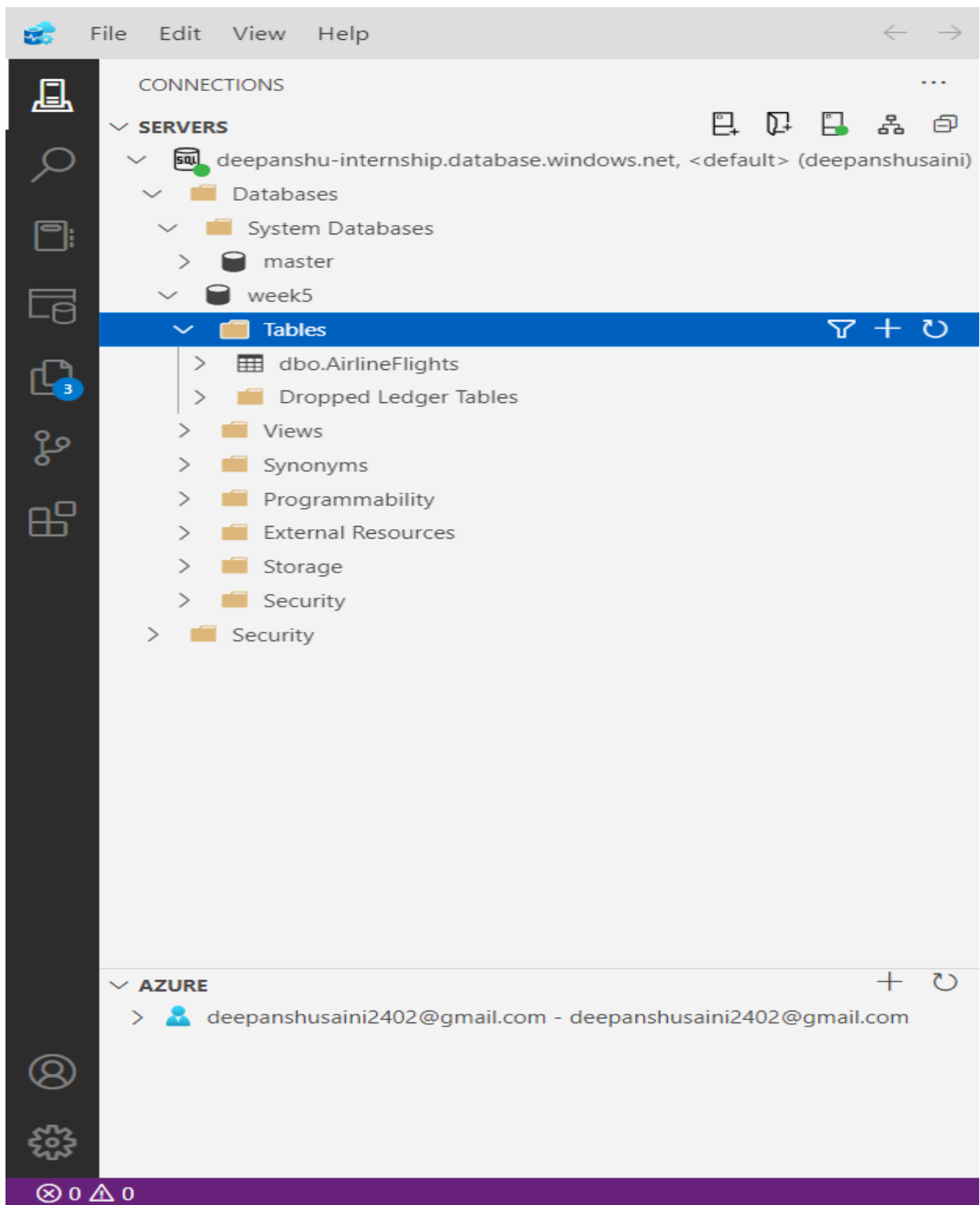
ASSIGNMENT

WEEK 5

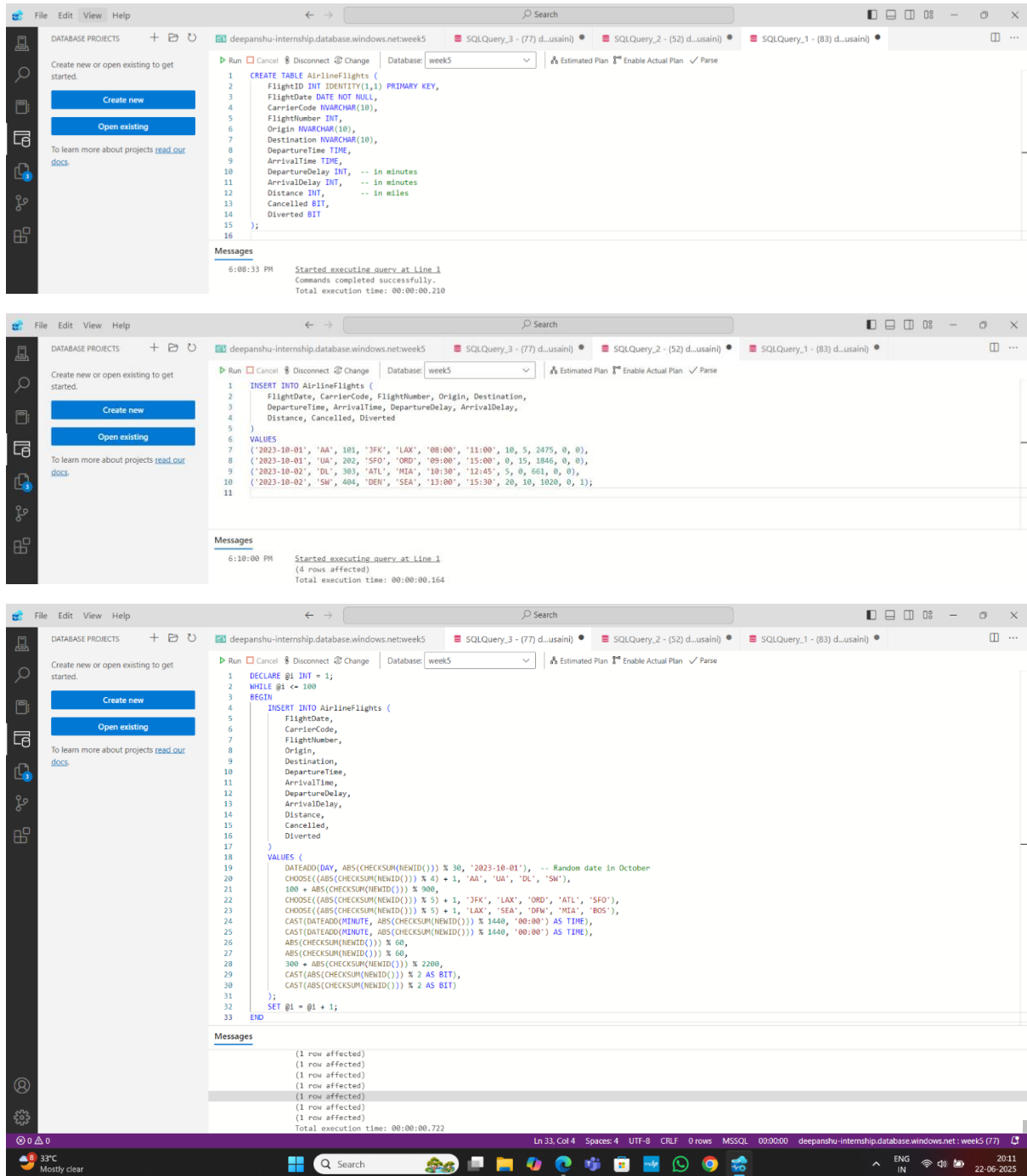
PROBLEM STATEMENT: This set of tasks focuses on data integration and pipeline automation. It involves copying data from a database into multiple file formats such as CSV, Parquet, and Avro for storage or further processing. The pipeline is automated using scheduled and event-based triggers to ensure timely and efficient execution. Additionally, it includes copying all tables from one database to another, as well as selectively transferring specific tables and columns based on business requirements, enabling fine-grained control over data migration and transformation.

STEP 1: Database Deployment using Azure Data Studio over Azure SQL Database:

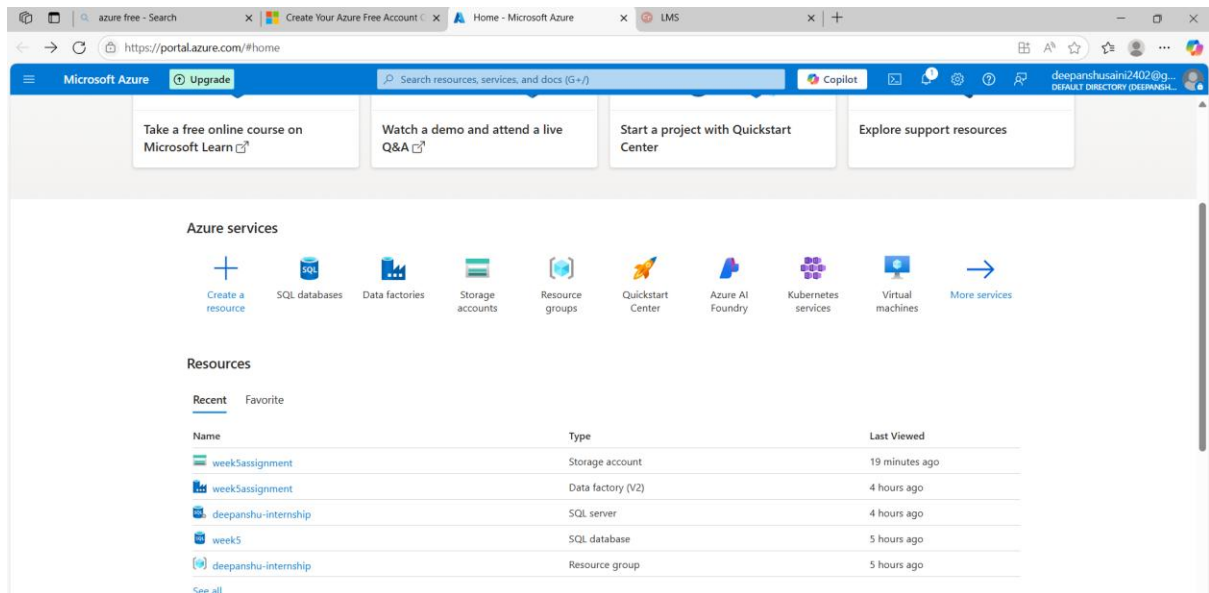
- Connection with Azure SQL Database



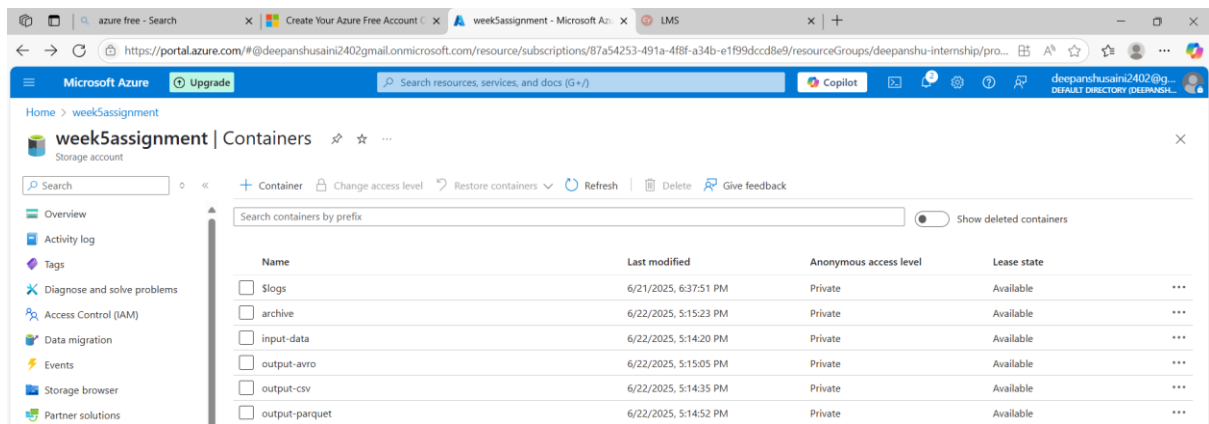
- Queries for creating table and inserting values into it.



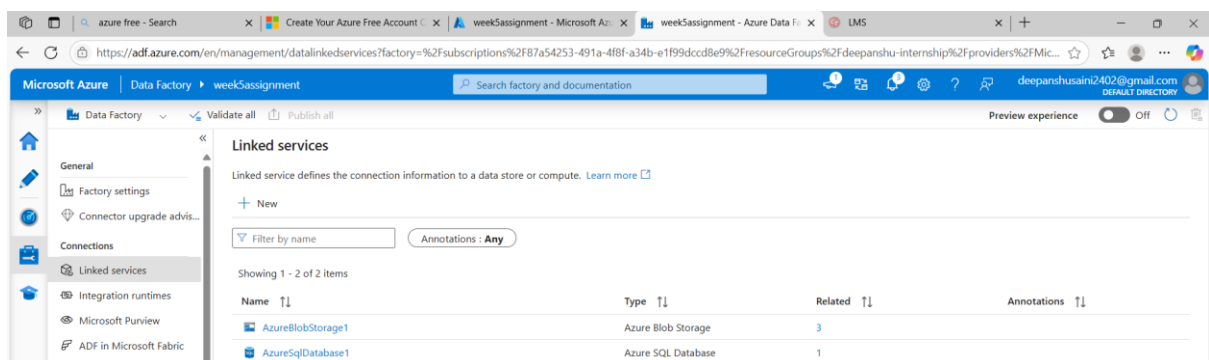
STEP 2: Creating Resources on Azure Portal



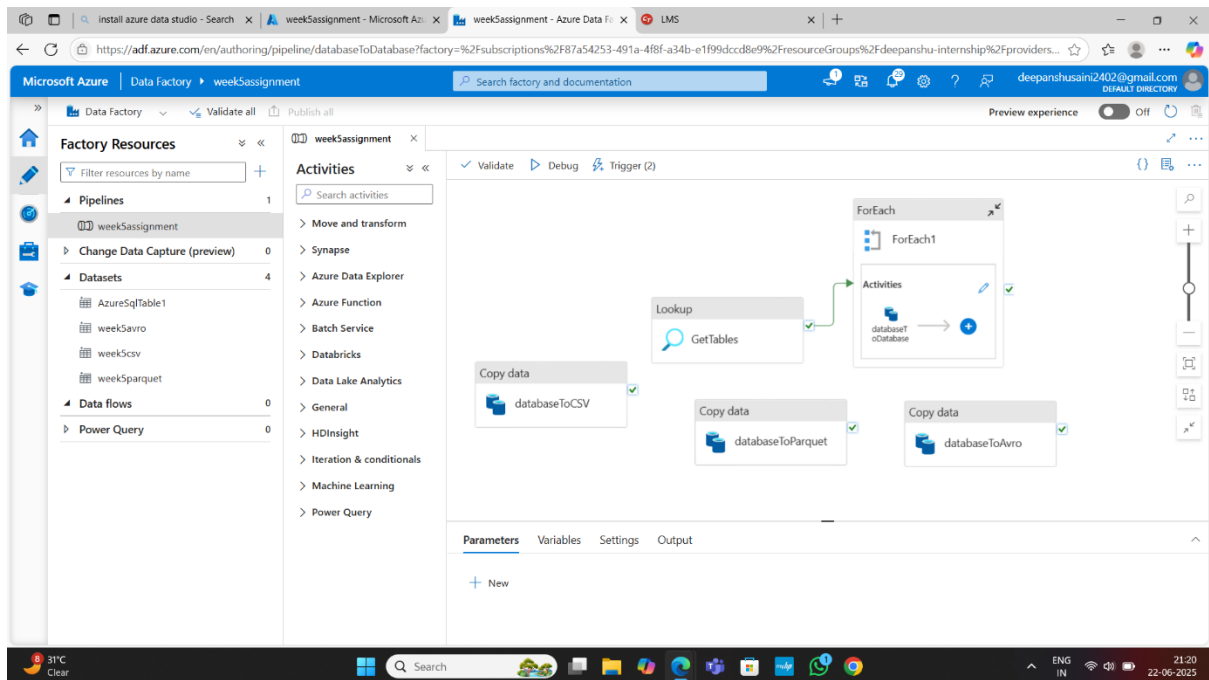
STEP 3: Creating Containers inside Storage Account.



STEP 4: Linking Services with Azure Data Factory.



STEP 5: Creating Pipeline



The pipeline include:

- databaseToCSV (Copy data activity)
- databaseToParquet (Copy data activity)
- databaseToAvro (Copy data activity)
- GetTables (Lookup activity – parent) -> For each (Child) -> DatabaseToDatabase (Copy data activity)
- scheduleRunning (Schedule based Trigger)
- Event based Trigger

Add triggers

Choose trigger...

Trigger	Type	Parameters	
tumblingWindowTri...	Tumbling window	0	×
scheduledRunning	Schedule	0	×